REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
18-03-2015	Final Report	1-Sep-2011 - 31-Oct-2014

4. TITLE AND SUBTITLE
Final Report: Creating an Online Laboratory

5a. CONTRACT NUMBER
W911NF-11-1-0392

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
David Lazer, Waleed Melies, Jason Radford

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Northeastern University
360 Huntington Ave
960 RP
Boston, MA 02115 -5005

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
60315-LS.7

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Existing online experimental research uses customized software integrating with specific, predefined subject pools such as Mechanical Turk to perform single experiments. The weakness of this decentralized, ad hoc approach to online experimentation is a continued dependence on specific subject pools, experiments that remain difficult to replicate or transfer to other researchers, and a continuous re-inventing the wheel as every social scientist must solve the same problems such as recruiting and managing large numbers of participants. Since 2011, we developed a robust, scalable, and flexible platform called Volunteer Science (http://volunteerscience.com) for designing

15. SUBJECT TERMS
Online Research, Networks, Experimentation, Computational Social Science

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			David Lazer
UU	UU	UU	UU		19b. TELEPHONE NUMBER
					617-373-8013

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: Creating an Online Laboratory

ABSTRACT

Existing online experimental research uses customized software integrating with specific, predefined subject pools such as Mechanical Turk to perform single experiments. The weakness of this decentralized, ad hoc approach to online experimentation is a continued dependence on specific subject pools, experiments that remain difficult to replicate or transfer to other researchers, and a continuous re-inventing the wheel as every social scientist must solve the same problems such as recruiting and managing large numbers of participants. Since 2011, we developed a robust, scalable, and flexible platform called Volunteer Science (http://volunteerscience.com) for designing, testing, and running experiments online. This platform addresses numerous challenges with online research and laboratory experimentation in general, making it easy to implement, modify, elaborate on, validate, replicate, and audit research. In addition, we built a diverse set of subject recruitment strategies including integrating with social media, Amazon's Mechanical Turk, and class-based research participation. Finally, we implemented and tested six experimental platforms from across disciplines to perform our own research and validate the platform.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Lazer, David. "Volunteer Science: A Crowd Sourced Platform for Studying Human Behavior." Conference on Digital Experimentation. Cambridge, MA. October 2014.

Radford, Jason, David Lazer, Brooke Foucault-Welles, Brian Keegan, Christoph Riedl, M. Scott Poole, and Jefferson Hoye. “Conducting Massively-Open Online Social Experiments with Volunteer Science.” Citizen+X at AAAI Conference on Human Computation and Crowdsourcing. Pittsburgh, PA. November 2014.

Number of Presentations: 2.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
03/16/2015	5.00	Ceyhun Karbeyaz, David Lazer, Waleed Meleis, Alan Mislove, Skyler Place. A web-based laboratory for computational social science, Research, Innovation, and Scholarship Expo at Northeastern University. 01-APR-12, . : ,

TOTAL: **1**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
10/10/2014	4.00	Jason Radford, Brian Keegan, Jefferson Hoye, Ceyhun Karbeyaz, Waleed Meleis, David Lazer. Volunteer Science: A Web Laboratory for Massively Open Online Social Experiments, Human Computation - CitizenX Workshop. 02-NOV-14, . : ,

TOTAL: **1**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
03/16/2015	6.00	Jason Radford , Brian Keegan, Jefferson Hoye, Ceyhun Karbeyaz, Katherine Ognyanova, Brooke Foucault Welles, Waleed Meleis , David Lazer. Validating a Citizen Science Model for Online Social and Behavioral Experiments Using Volunteer Science, Social Psychology Quarterly (03 2015)
03/16/2015	3.00	Brian Keegan, Jason Radford, Brooke Foucault Welles, Christoph Riedl, Jefferson Hoye, Ceyhun Karbeyaz, Waleed Meleis, David Lazer. Conducting Massively Open Online Social Experiments with Volunteer Science, (03 2015)

TOTAL: **2**

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Ceyhun Karbeyaz	0.25	
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Brian Keegan	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
David Lazer	0.10	Yes
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME	
Total Number:	

Names of personnel receiving PHDs

NAME	
Ceyhun Karbeyaz	
Total Number:	**1**

Names of other research staff

NAME	PERCENT_SUPPORTED
Jefferson Hoye	1.00
FTE Equivalent:	**1.00**
Total Number:	**1**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Attachment

Technology Transfer

Creating an Online Laboratory

David Lazer, Waleed Melies, Jason Radford

Statement of the Problem Studied

Performing experimental research on social networks is challenging for the dearth of experimental research on networks, the difficulty developing and managing the experiments, and challenges with existing subject pools traditionally used for network experiments. While experimental research on group processes, individual decision making, and cognition have long histories, only recently have researchers returned to the question of how individuals perform in networks and how the structure of social networks affects individual and collective behavior. One reason for this is that network experiments require large numbers of participants in laboratory setting which require overlapping but incongruous sets of individuals to interact with one another. They are thus particularly difficult to manage. The advent of methods for online experimentation have greatly reduced these barriers, but current approaches to online network research remain fragmentary and tied to problematic sample pools.

Review of Network Experimentation

How do we think together? And how do the connections within a group affect how individuals in that group think together? Despite the apparent breadth of these issues, there is surprisingly little research that addresses these questions. There is a large literature on groups, and the determinants of group performance (see review by Hackman and Katz 2011), including a significant amount on the micro to macro connection between individuals and groups (Sunstein 2009). There is a relatively small, if growing, literature, however, on group networks (see review by Katz et al. 2004), and only a tiny subset of that is about how interconnectedness affects how people think together.

We define "thinking together" as the emergent nature of the set of answers and conceptualizations of a problem that confronts multiple individuals. How to solve a puzzle, the shortest route to work, what restaurant to have dinner at, are all problems that have an individual locus for decision making. They are, however, collectively solved, in that we talk to, learn from, and observe what other people do. Our individual instincts in how to solve a problem must surely be a mix of solving the problem based on our particular capacities, while incorporating the information available in our network. The objective of the proposed research is to understand, (1) at the micro level, how individual problem solving is affected by social cues, and (2) at the macro level, what are the collective, emergent properties of those individual behaviors. The resulting insights have potentially broad implications for understanding collective cognition at various scales, from small groups to societies.

This research builds on agent-based modeling on these issues by the PI that focused on one of the key ways that the network affects collective problem solving, which is in balancing exploration and exploitation (Lazer & Friedman 2007; March 1991). Exploration is defined as the development of new solutions to a problem, and exploitation the use of known solutions. The essential decision-making conundrum is that exploration may come at the opportunity cost of taking advantage of current best practices. Lazer and Friedman argue that the structure of a group's network determines the balance between exploration and exploitation. The more efficient the network of a group, the more rapidly it converges on known best solutions. Convergence of known best solutions, however, reduces diversity, which, in turn, reduces performance (Page 2008) because of the reduction in exploration. That is, efficient communication reduces long run performance.

A key shortcoming of this research is that it builds on a very narrow body of experimental research. This is an approach to the study of social networks that largely fell into disuse for many decades (Katz et al. 2004), with a vigorous vein of experimental research on small (usually < 10) group networks conducted in the 1950s (Bavelas 1950; Guetzkow & Simon 1955), but largely dying out by the 1960s. For the types of questions we wish to explore, such as examination of the impact of network topology, larger groups would be useful, which puts a strain on the standard human behavioral laboratories. Our experience with traditional laboratory experiments highlights the need for an alternative model. In our own experiments, after scaling down the group size and reducing the experimental parameters, we still exhausted the entire subject pools of the two laboratories we used, spent a large amount of resources on subject costs, and took an entire year to run trials.

Online Experimentation

In part based on these experiences, and in part due to the emerging literature on online experiments, it is clear that creating an online group network research laboratory is necessary in order to pursue these research questions. Conducting experiments online offers enormous potential for studying the implications of different network structure. Watts and collaborators (Salganik et al. 2006), Centola (Centola 2010), and Goldstone (Goldstone et al. 2008), among others, have broken critical ground in conducting experimental online research. However, existing online experimental research uses customized software integrating with specific, pre-defined subject pools such as Amazon's Mechanical Turk (AMT) to perform single experiments. The weakness of this decentralized, ad hoc approach is a continued dependence on specific subject pools, experiments that remain difficult to replicate or transfer to other researchers, and a continuous re-inventing the wheel as every social scientist must solve the same programmatic issues like subject-matching, consent management, and data retention.

The development of online experiments and crowd-based scientific research more generally has reduced the costs of research while broadening the number of people who can participate and the number of ways to participate. Current approaches to online experimentation in the social sciences are largely research-specific and tied to individual subject pools such as workers on AMT and similar platforms, online samples such as GfK (formerly, Knowledge Networks), or computer-based experiments run in physical laboratories. This ad hoc state of the field generates large start-up costs for online experimentation, makes the field dependent on specific subject pools, and perpetuates highly stylized experiments which are difficult to replicate or borrow from.

Other initiatives in online experimentation and citizen science demonstrate the need for a more open, common platform based on a citizen science model. For large-scale, online social experiments, programs like Time-sharing Experiments in Social Sciences (TESS) and Project Implicit have proven to be powerful engines for collecting data. However, they provide limited collective benefits and remain narrow in scope. TESS provides a very narrow set of experimental templates and relies on a very expensive subject pool. Project Implicit relies on a volunteer sample, but is usable only for a very narrow array of scientific questions. Citizen science programs in other fields like Galaxy Zoo and Fold.it demonstrate the willingness of individuals to participate in scientific research and the benefits of having a publically open, participatory system. We sought to combine the strengths of TESS and Project Implicit by creating systems which facilitate developing and deploying online experiments with the strengths of citizen science projects which generate communities of thousands of participants from across the globe.

Summary of the Most Important Results

For this proposal, we built the Volunteer Science research platform, implemented six experimental paradigms, and presented the results of our research to audiences at seven conferences and wrote two papers to be submitted this Spring. Funding from ARO was used to design the Volunteer Science website, develop the experimental Application Programming Interface (API) and documentation, and integrate with external resources like AMT, Facebook, and the survey software Qualtrics. We also created working versions of the travelling salesperson problem, prisoners' dilemma, public goods game, ultimatum game, word ladders, and reaction time experiments. These experiments (along with others) culminated in two papers, one focusing on the design of the platform and the other on the results of research testing validity of the platform and its volunteer-based methods.

The Volunteer Science Platform

Volunteer Science was created to simplify design, make research easy to share and iterate on, and collectivize subject management processes. "Volunteer Science" is both a website and development platform for deploying web-based single-user and multi-user behavioral experiments. This site is built on top of free open source development tools including Django, Bootstrap, HTML5, and JavaScript and uses Amazon Web Services (AWS) to provide high-performance and a scalable on-demand hosting. These web-based technologies are agnostic to platform (Windows, Mac, Linux, Internet Explorer, Firefox, Chrome) and also support mobile devices. It is hosted on AWS which allows the platform to scale up and down to meet changing demand rather than requiring researchers to purchase, maintain, and upgrade specialized hardware.

The technical core of Volunteer Science for researchers is the body of experiment templates and API codes. The API makes complex experimental protocol like subject matching, consent, and real-time communication easy to use. In developing fully functional implementations of experiments like prisoner's dilemma and the ultimatum game, we also developed a template for bargaining games in general, enabling researchers to directly implement other paradigms like the tragedy of the commons, public goods games, or any Nash bargaining game. Finally, included in this API are easy-to-use codes for accessing external resources essential to online experimentation like Facebook, Twitter, Qualtrics, and AMT. This enables researchers to recruit subjects broadly and access the broader ecology of data available online.

Experiments and Research on Volunteer Science

We originally built Volunteer Science to conduct a set of social and behavioral science experiments requiring synchronous communication among subjects for researchers at three universities. The model for these experiments was to define a meaningful task and have people perform the task within networks of others performing the task. Thus, we built five problems which could be embedded in networks: the traveling salesperson, prisoner's dilemma, public goods, ultimatum game, and word ladders. The sixth, reaction time experiments, was built to expand the research into broader disciplines for the purpose of validating the platform.

Traveling Salesperson. We gave every individual an identical Traveling Salesperson Problem (TSP) to solve, a canonical computer science problem that involves identifying the shortest itinerary for a hypothetical salesman traveling among a set of cities. We manipulated the types of information and connections to others' solutions and our preliminary finding confirms the essential explore/exploit tradeoff: more communication yielded fewer good answers, but faster dissemination of what good answers were found.

Prisoner's Dilemma. The classic bargaining game was embedded in networks of individuals to determine whether or not players in more central positions in a network would have more influence over the strategies of players.

Ultimatum Game. The one-sided bargaining game was created to test the transitivity of bargaining to determine whether someone who received a high or low value ultimatum would transfer that to a third person.

Public Goods. The public goods game tests the ability for individuals to contribute to a collective pool in order to reap larger rewards while preventing free riding. In the network version we created, we test whether or not individuals perform differently when they can see the contribution of their neighbors which may be one other person in the group, the whole group, or everything in between.

Word Ladders. The word ladder game is a puzzle wherein individuals are given one word like "dog" and, by changing one letter at a time, get the word "cat" (e.g. dog-cog-cot-cat). By being exposed to others solving these problems, we believe individuals will exploit others solutions, leading to more rapid convergence to the best solution, but with fewer solutions generated than under the non-exposure condition.

Reaction Time Experiments. We developed versions of the Stroop (MacLeod 1991) and flanker tests (Erikson 1995) which detect very slight delays in individual reaction times based on simple manipulations. Replicating these tests provides support for using Volunteer Science to perform time-sensitive research and provided for the template for many other common psychological experiments and measures pertaining to memory, perception, and cognition.

Scientific Products

The results of these three years of funding have been two papers currently being written and which we plan to submit for publication this Spring. The first paper documents the design of the Volunteer Science system and the advantages and disadvantages of our model. The purpose of this paper is to bring transparency to the system, document the decisions we made, and offer guidance to other researchers performing online experiments or working with volunteers. The second paper will focus on how we validated the platform by replicating a series of canonical experiments and findings. We included a broad variety of experiments and measures from psychology, economics, and problem solving to provide robust tests of standard effects. The results show that our users participate in good faith, are surprisingly diverse, and can be used to study a number of core phenomena of interest to social scientists.

Conclusion

Volunteer Science enables researchers to quickly and easily build online experiments by giving them access to existing templates, a range of important features, diverse subject pools, and integrates directly with other means of online research. We have demonstrated the power of the platform by implementing a range of canonical experiments, problems, and puzzles and tested how users behave in social networks. And, now that we've validated the platform and we are starting to increase the number of volunteers participating in our experiments, the results of our research should begin to emerge and the platform should expand to include a broader variety of collaborators and experiments.

Bibliography

Bavelas, A., 1950. Communication Patterns in Task-Oriented Groups. *Journal of the Acoustical Society of America*, 22, pp.725–730.

Centola, D., 2010. The Spread of Behavior in an Online Social Network Experiment. *science*, 329(5996), p.1194.

Eriksen, Charles W. 1995. “The Flankers Task and Response Competition: A Useful Tool for Investigating a Variety of Cognitive Problems.” *Visual Cognition* 2 (2-3): 101–18.

Goldstone, R.L., Roberts, M.E. & Gureckis, T.M., 2008. Emergent Processes in Group Behavior. *Current Directions in Psychological Science*, 17(1), p.10.

Guetzkow, H. & Simon, H.A., 1955. The impact of certain Communication Nets upon Organization and Performance in Task-Oriented Groups. *Management Science*, 1(3), pp.233–250.

Hackman, J.R. & Oldham, G.R., 1976. Motivation through the Design of Work: Test of a Theory. *Organizational behavior and human performance*, 16(2), pp.250–279.

Katz, N. et al., 2004. Network Theory and Small Groups. *Small Group Research*, 35(3), p.307.

Lazer, D. & Friedman, A., 2007. The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52(4), pp.667–694.

MacLeod, C.M. 1991. “Half a Century of Research on the Stroop Effect: An Integrative Review.” *Psychological Bulletin* 109 (2): 163–203.

March, J.G., 1991. Exploration and Exploitation in Organizational Learning. *Organization science*, 2(1), pp.71–87.

Page, S.E., 2008. *The Difference: How the Power of Diversity creates better Groups, Firms, Schools, and Societies*, Princeton: Princeton University Press.

Salganik, M. J., and Watts, D.J. 2008. “Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market.” *Social Psychology Quarterly* 71 (4): 338–55.

Sunstein, C.R., 2009. *Going to Extremes: How Like Minds Unite and Divide*, Oxford University Press, USA.

Watts, D.J. & Strogatz, S.H., 1998. Collective Dynamics of ‘Small-World’ Networks. *Nature*, 393(6684), pp.440–442.